WHAT IS CLAIMED IS:

1   1.   A system for duplicating documents of disparate types to provide a searching

2   mechanism, said system comprising:

3   storage means for storing an original file containing one original document, said

4   original file being in a first format;

5   processing means when said original document is a component document of a

6   compound document for extracting said component document from said compound document and

7   storing said component document in said original file; and

8   conversion means for converting said original file to a canonical format to produce

9   a duplicate file containing a duplicate document of said original document.

1   2.   A system for securely duplicating documents of disparate types to provide a

2   cryptographically secure link between a duplicate document and an original document, said system

3   comprising:

4   storage means for storing an original file containing one original document, said

5   original file being in a first format;

6   conversion means for converting said original file to a canonical format to produce

7   a duplicate file containing a duplicate document of said original document;

8   embedding means for embedding notary data associated with said original file into

9   said canonical format of said duplicate file, said notary data being capable of authenticating said

10   original file; and

11   indexing means for inserting a sequence number including part of said notary data

12   into said duplicate file.

1   3.   The system of Claim 2, wherein said original document is a component document

2   of a compound document, and further comprising:

3   processing means for extracting said component document from said compound

4   document and storing said component document in said original file.

1   4.   The system of Claim 3, wherein said storage means has a directory structure, said

2   component document being stored in said storage means under a directory associated with said

3   compound document.

1   5.   The system of Claim 2, wherein said canonical format is the Portable Document

2   Format (PDF).

1   6.   The system of Claim 5, wherein said duplicate document has a Root Dictionary

2   field therein, and further comprising:

3   creation means for creating a new Dictionary associated with said system; and

4   insertion means for storing said new Dictionary in said Root Dictionary field and

5   inserting said notary data into said new Dictionary.

1   7.   The system of Claim 2, wherein said notary data includes at least a digital

2   fingerprint associated with said original file, a timestamp indicating the time said digital

3   fingerprint was obtained and an identifier, said sequence number including said identifier.

1   8.   The system of Claim 2, wherein said sequence number is inserted as a footer on

2   the bottom of each page of said duplicate document.

1   9.   The system of Claim 2, further comprising:

2   additional storage means for storing said duplicate file, said additional storage

3 means having a directory format identical to the directory format of said storage means storing

4 said original file.

1   10.   The system of Claim 2, further comprising:

2   additional embedding means for embedding notary data associated with said

3 duplicate file into said canonical format of said duplicate file, said notary data being capable of

4 authenticating said duplicate file.

1   11.   The system of Claim 2, further comprising:

2   a log file configured to map a sequenced filename associated with said duplicate

3 file to a filename associated with said original file.